# Mixed-state Pauli-channel parameter estimation

David Collins[*]

*Department of Physical and Environmental Sciences, Colorado Mesa University, Grand Junction, Colorado 81501, USA*
(Received 30 August 2012; published 4 March 2013)

The accuracy of any physical scheme used to estimate the parameter describing the strength of a single qubit Pauli channel can be quantified using standard techniques from quantum estimation theory. It is known that the optimal estimation scheme, with $m$ channel invocations, uses initial states for the systems which are pure and unentangled and provides an uncertainty of $O(1/\sqrt{m})$. This protocol is analogous to a classical repetition and averaging scheme. We consider estimation schemes where the initial states available are not pure and compare a protocol involving quantum correlated states to independent state protocols analogous to classical repetition schemes. We show, that unlike the pure state case, the quantum correlated state protocol can yield greater estimation accuracy than any independent state protocol. We show that these gains persist even when the system states are separable and, in some cases, when quantum discord is absent after channel invocation. We describe the relevance of these protocols to nuclear magnetic resonance measurements.

PACS number(s): 03.67.−a, 03.65.Ta, 03.65.Ud

## I. INTRODUCTION

Quantum estimation considers situations in which a quantum system or evolution operation depends on one or more parameters, the values of which are to be estimated as accurately as possible. Estimating such parameters can only be accomplished by subjecting physical systems to physical procedures. Consequently the accuracy of estimates will be governed by the laws of quantum physics. This realization has resulted in substantial study of constraints and quantum advantages available for estimation in the context of quantum theory and has yielded results with important implications for quantum metrology [1–13].

In general, any quantum estimation procedure considers a quantum operation of known form but which depends on one or more parameters. The focus of this article is the Pauli operation (or channel), $\hat{\Gamma}(\lambda)$, which acts on a single qubit. This transforms a system, initially in the state $\hat{\rho}$, via

$$\hat{\rho} \overset{\hat{\Gamma}(\lambda)}{\mapsto} \hat{\rho}_\mathrm{f}(\lambda) := (1 - \lambda)\hat{\rho} + \lambda\hat{\sigma}_n\hat{\rho}\hat{\sigma}_n \quad (1)$$

where $0 \leqslant \lambda \leqslant 1$ and $\hat{\sigma}_n$ is a Pauli operator with $n$ represent a direction in space. For example, $n = x$ represents the bit-flip operation and $n = z$ is the phase-flip operation. The assumption is that the direction $n$ is known but the parameter $\lambda$ is unknown. The task is to estimate $\lambda$ with minimal uses of physical resources, such as qubits, number of applications of the channel or ancillary quantum operations.

Channels of this type are important for the following reasons. First, the three channels for which $n = x$, $y$, and $z$ are sufficient to generate the types of errors that are considered in quantum information processing [14]. Estimating their strengths via the parameters will be relevant for practical quantum information processing protocols.

Second, specific examples of such channels appear in contexts such as nuclear magnetic resonance (NMR), in which the phase-flip operation generates the dephasing (or $T_2$) processes [15,16]. Here the density operator of a single

spin-1/2 system evolves with time $t$ as

$$\hat{\rho}(t) = \begin{pmatrix} \rho_{00} & \rho_{01}e^{-t/T_2} \\ \rho_{10}e^{-t/T_2} & \rho_{11} \end{pmatrix} \quad (2)$$

where $T_2$ is a constant with units of time. This is related to the phase-flip parameter by

$$\lambda = (1 - e^{-t/T_2})/2. \quad (3)$$

Measuring or estimating the phase-flip parameter is thus equivalent to measuring $T_2$.

Third, quantum parameter estimation can illustrate fundamental differences between classical and quantum worlds. For example, consider two methods of estimating the parameter governing unitary phase shifts [8]. Applying the unitary phase operation once to each of $n$ independent quantum systems, yields an optimal uncertainty of $O(1/\sqrt{n})$. This is analogous to a classical repetition and averaging procedure. A method which applies the unitary phase operation once to each of $n$ quantum systems which are initially in a particular entangled state yields an uncertainty of $O(1/n)$. This is optimal for this type of operation [8,17,18]. Evidently the resource of entanglement, occurring in quantum systems, has yielded an advantage here.

Parameter estimation for Pauli channels has been considered in the context of finding an optimal scheme using all possible input states of the quantum systems involved [7,19–23]. The results, reviewed in Sec. III, are that the optimal accuracy can be attained when the channel is invoked multiple times on independent and unentangled systems, each of which is initially in a pure state [19]. The manifestly quantum resource of entanglement has no role to play in this circumstance. However, in situations such as solution state NMR, initial states are highly mixed. The question remains as to whether approaches similar to those involving entangled states can provide advantages over the independent channel invocation approach when the only initial states available are mixed. This variant of quantum parameter estimation has been investigated for unitary operations [24–27] where advantages of the type using entangled pure states survived with mixed, correlated initial states.

————————
[*]dacollin@coloradomesa.edu

In this article we consider the question of Pauli channel parameter estimation when the initial states of the available quantum systems are mixed. We aim to compare the accuracy of an independent channel use approach to one where correlated or entangled states are used. Our goal is to determine whether correlated state approaches can yield gains for Pauli channel parameter estimation, as they do for mixed state phase shift estimation [26]. This could be important for Pauli channels in solution state NMR and other situations where the initial states are inherently mixed. Particularly, for NMR it is not practical to purify the state by a process of measurement followed by selection based on measurement outcomes; the full range of available initial states does not include pure states. Our approach is similar to that of Modi *et al.* [26] but the results are subtly different, particularly in terms of the relationship between the presence of entanglement or discord when improved accuracy is attained.

We shall focus on the phase-flip channel but the results hold for any Pauli channel. Section II reviews the general framework for quantum parameter estimation and general bounds for phase-flip parameter estimation. Section III describes initial states available and gives the optimal independent channel use estimation accuracy; this provides a "classical" benchmark against which other schemes can be compared. Section IV outlines a particular approach that uses correlated quantum states and gives central new results which form the foundation our the rest of the article. Sections V and VI describe two cases of the main result with notably different outcomes. Finally Sec. VII discusses the presence of correlations such as entanglement or quantum discord in our estimation schemes.

## II. QUANTUM PARAMETER ESTIMATION

Typically a physical quantum estimation procedure commences with various quantum systems prepared in known initial states. The parameter dependent operation is applied multiple times to some of these systems, possibly interspersed with ancillary parameter-independent quantum operations. The physical procedure terminates with measurements on the individual quantum systems. The parameter is inferred by data processing from the measurement outcomes. Figure 1 illustrates a generic quantum estimation scheme which uses $m$ invocations of the operation and $n$ quantum systems.
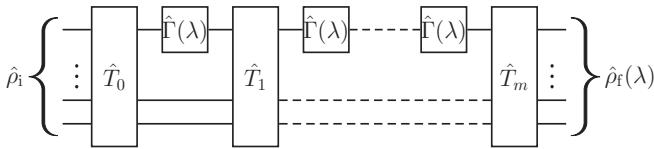


FIG. 1. Schematic of a quantum estimation procedure using $n$ quantum systems, each represented by a horizontal line. This procedure uses $m$ invocations of the parameter-dependent operation $\hat{\Gamma}(\lambda)$, interspersed with parameter-independent unitary transformations $\hat{T}_1, \ldots, \hat{T}_m$. The initial state of all quantum systems is $\hat{\rho}_i$ and the premeasurement state $\hat{\rho}_f(\lambda)$. Measurements on individual systems follow the final unitary $\hat{T}_m$.

The general probabilistic nature of the outcomes of measurements on quantum systems means that the parameter inferences will fluctuate about the true parameter value. A key goal of any parameter estimation scheme is to minimize these fluctuations.

Various estimation schemes use differing resources (e.g., number of qubits, number of channel invocations, number of two qubit entangling unitaries) and a fair comparison requires precise specification of the resource costs. A common assessment of the resource cost uses a black-box query count within a quantum circuit [8,17,18,28] such as illustrated in Fig. 1. In this approach, used in this article, the only relevant resource cost is the number of times, $m$, that the phase-flip operation appears in the quantum circuit for the protocol. We will only compare schemes that use the same number of channel invocations.

In the scheme of Fig. 1, $\hat{\rho}_i$ can be assumed to be a product of states for the individual qubits since an potential entangled state can be generated from this via an appropriate unitary, which can be subsumed into $\hat{T}_0$. Given this, estimation schemes can broadly be grouped into two classes.

In the first, called an *independent channel use protocol,* the channel is applied once to each of $m$ qubits, which are initially in a product state. This is followed by measurements. In the context of Fig. 1 this means that all of $\{\hat{T}_1, \ldots, \hat{T}_m\}$ are nonentangling among the qubits (a nonentangling unitary transforms any state which is a product of states for individual quantum systems to another such product state). In this protocol $\hat{T}_0$ is a product of single qubit unitaries while $\hat{T}_1, \ldots, \hat{T}_m$ are products of single qubit unitaries sandwiched between two qubit swap gates. Throughout the procedure of this protocol the state of the entire system is always a product of states for the individual qubits and there are no quantum correlations present. The independent channel use protocol is analogous to the classical strategy of invoking the physical operation independently on multiple physical systems; this generally enhances estimation accuracy.

In the second approach, called a *correlated state protocol* and only available to quantum systems, at least one of $\{\hat{T}_1, \ldots, \hat{T}_m\}$ is an entangling unitary and, at some point during the procedure, the system will not be in a product state.

The task of schemes falling into either class is to find admissible unitaries $\{\hat{T}_1, \ldots, \hat{T}_m\}$ that maximize the estimation for fixed resource costs. The question is whether there is some correlated state protocol that provides better estimation accuracy that any independent state protocol. If so, it would indicate that some quantum resources, possibly generated by an entangling unitary, have been harnessed so as to outperform any estimation protocol merely analogous to the classical strategy of repetition, as employed by an independent channel use protocol.

### A. Quantum Fisher information

Quantifying the accuracy of quantum estimation procedures is done using the following general scheme. Denoting the outcome of the measurement on the $j$th system by $x_j$, the collective outcomes of the physical procedure of Fig. 1 can be represented by $\{x_1, \ldots, x_n\}$. The parameter is estimated by applying these an *estimator function*, $\lambda_{est}$, resulting in

the estimate, $\tilde{\lambda} = \lambda_{\text{est}}(x_1, \ldots, x_n)$. As a result of statistical fluctuations in measurement outcomes, inherent in quantum systems, the same estimator will yield possibly different results from one run to another which uses an identical sequence of quantum operations. The statistical accuracy of the estimate will therefore rely on the choice of estimator together with the probability distribution for the measurement outcomes. One means of quantifying the accuracy of this estimate uses the mean square error,

$$\text{mse}(\tilde{\lambda}) := \langle (\tilde{\lambda} - \lambda)^2 \rangle \qquad (4)$$

where the angle brackets indicate the mean over all possible measurement outcomes. The possible estimators consider in this article will be *unbiased estimators*, i.e., $\langle \tilde{\lambda} \rangle = \lambda$, and here $\text{mse}(\tilde{\lambda}) = \text{var}(\tilde{\lambda})$, where the variance in the estimate is

$$\text{var}(\tilde{\lambda}) := \langle (\tilde{\lambda} - \langle \tilde{\lambda} \rangle)^2 \rangle. \qquad (5)$$

The variance is the fundamental entity which quantifies the accuracy of the estimate. A fundamental result [9,11,29] is that, *regardless of the estimator function, the variance is bounded from below via the Cramér-Rao bound*

$$\text{var}(\tilde{\lambda}) \geqslant \frac{1}{F(\lambda)} \qquad (6)$$

where

$$F(\lambda) := \int \left[ \frac{\partial \ln p(\vec{x}|\lambda)}{\partial \lambda} \right]^2 p(\vec{x}|\lambda) \, dx_1 \ldots dx_n \qquad (7)$$

is the Fisher information associated with the joint probability distribution (conditional upon $\lambda$) for the measurement outcomes, $p(\vec{x}|\lambda)$. Here $\vec{x}$ is an abbreviation for $x_1, \ldots, x_n$. If each measurement outcome is one of a set of discrete possibilities, the integral is replaced by a summation.

It can be shown that under general conditions, there is always an estimator [29], which asymptotically saturates the bound of Eq. (6). Thus in classical parameter estimation the Fisher information quantifies accuracy of the estimate and when comparing estimation schemes, that with the greater Fisher information is more successful than any competitor with a lower Fisher information.

Given a particular estimation procedure with Fisher information, $F_1(\lambda)$, an obvious and conventional classical strategy for improvement is to repeat the identical procedure multiple times independently. Mathematically this is equivalent to sampling the same probability distribution in an independent and identical fashion. If this is done $m$ times then the resulting overall Fisher information is easily shown [11] to be $m F_1(\lambda)$ and thus, in this scenario,

$$\text{var}(\tilde{\lambda}) \geqslant \frac{1}{m F_1(\lambda)} \qquad (8)$$

where $F_1(\lambda)$ is the Fisher information associated with a single application of the estimation procedure. This governs the typical behavior as a function of $m$ in classical parameter estimation. For example, in the context of parameter estimation in classical physical systems, the same physical procedure can be applied to $m$ identical systems. The resulting mean square error is reduced by a factor of $m$ and the uncertainty by a factor of $\sqrt{m}$.

For parameters pertaining to the evolution of quantum systems, the probability of measurement outcomes is determined by (i) the choice of measurement and (ii) the choice of initial state $\hat{\rho}_{\text{i}}$, which determines the final state $\hat{\rho}_{\text{f}}(\lambda)$. Therefore, the classical Fisher information for any given final state depends on the choice of the measurement. However, considering all possible quantum measurement procedures on the system and any possible ancillary quantum systems shows that the classical Fisher information satisfies the *quantum Cramér-Rao bound*,

$$F(\lambda) \leqslant H(\lambda) \qquad (9)$$

where the quantum Fisher information $H(\lambda)$ is *independent of the choice of measurement* [5,9,11]. The quantum Fisher information depends only on $\hat{\rho}_{\text{f}}(\lambda)$ and is calculated via [9]

$$H(\lambda) = \text{Tr}[\hat{\rho}_{\text{f}}(\lambda)\hat{L}^2(\lambda)] \qquad (10)$$

where the symmetric logarithmic derivative (SLD) or score operator, $\hat{L}(\lambda)$ is defined implicitly via

$$\frac{\partial \hat{\rho}_{\text{f}}(\lambda)}{\partial \lambda} = \frac{1}{2} [\hat{L}(\lambda)\hat{\rho}_{\text{f}}(\lambda) + \hat{\rho}_{\text{f}}(\lambda)\hat{L}(\lambda)]. \qquad (11)$$

A useful alternative to Eq. (10) is

$$\begin{aligned} H(\lambda) &= \text{Tr}[\hat{\rho}_{\text{f}}(\lambda)\hat{L}^2(\lambda)] \\ &= \frac{1}{2} \text{Tr}[\hat{L}(\lambda)\hat{\rho}_{\text{f}}(\lambda)\hat{L}(\lambda) + \hat{\rho}_{\text{f}}(\lambda)\hat{L}^2(\lambda)] \\ &= \text{Tr}\left[ \frac{\partial \hat{\rho}_{\text{f}}(\lambda)}{\partial \lambda} \hat{L}(\lambda) \right]. \end{aligned} \qquad (12)$$

There exist methods [9] for computing the SLD based on eigenvalue and eigenstate decompositions of $\hat{\rho}_{\text{f}}(\lambda)$ and conditions for the existence of measurements which saturate the bound of Eq. (9). In general there does exist a measurement procedure which asymptotically attains this bound [30]. Quantum Fisher information then serves as a measure of the possible accuracy of the estimation procedure. The actual measurement and data processing of the outcomes are irrelevant here and all that matters is the final state of the system, $\hat{\rho}_{\text{f}}(\lambda)$. The central task in any quantum parameter estimation scheme is to determine the initial state for the quantum system which maximizes the resulting quantum Fisher information and, if possible, the associated measurement which yields a classical Fisher information equal to the maximum quantum Fisher information. This reduces the task of designing quantum parameter estimation schemes to that of tailoring $\hat{\rho}_{\text{f}}(\lambda)$ so as to maximize the quantum Fisher information.

### B. Determining the SLD

Computing the SLD can be done by computing eigenstates and eigenvalues of $\hat{\rho}_{\text{f}}(\lambda)$ but this is generally difficult. However, for certain operators, there are methods for computing the SLD which do not require the eigenvalues and eigenstates.

*Proposition 1.* Let $\hat{A}$ be a diagonalizable operator on a two-dimensional vector space such that $\text{Tr}\hat{A} \neq 0$. Let

$\alpha := \mathrm{Tr}(\hat{A}^2) - (\mathrm{Tr}\hat{A})^2$. Then $\frac{\partial \hat{A}}{\partial \lambda} = (\hat{L}\hat{A} + \hat{A}\hat{L})/2$ where

$$\hat{L} = \frac{1}{\mathrm{Tr}\hat{A}} \left[ 2 \frac{\partial \hat{A}}{\partial \lambda} - \frac{\partial \ln(\mathrm{Tr}\hat{A})}{\partial \lambda} \hat{A} \right] \qquad (13)$$

if $\alpha = 0$ and

$$\hat{L} = \frac{1}{\mathrm{Tr}\hat{A}} \left[ 2 \frac{\partial \hat{A}}{\partial \lambda} - \frac{\partial \ln \alpha}{\partial \lambda} \hat{A} \right] + \frac{\partial}{\partial \lambda} [\ln \alpha - \ln(\mathrm{Tr}\hat{A})]\hat{I} \qquad (14)$$

if $\alpha \neq 0$.

*Proof.* Denote the eigenvalues of $\hat{A}$ by $a_1$ and $a_2$. The characteristic equation for $\hat{A}$ is $(\hat{A} - a_1\hat{I})(\hat{A} - a_2\hat{I}) = 0$. Expanding this gives

$$\hat{A}^2 - \mathrm{Tr}[\hat{A}]\hat{A} + a_1 a_2 \hat{I} = 0.$$

Further algebra yields

$$\hat{A} = \frac{1}{\mathrm{Tr}\hat{A}} \left( \hat{A}^2 - \frac{1}{2} \alpha \hat{I} \right) \qquad (15)$$

where we have used the fact that $\mathrm{Tr}\hat{A} \neq 0$. Thus

$$\frac{\partial \hat{A}}{\partial \lambda} = -\frac{1}{\mathrm{Tr}\hat{A}} \frac{\partial \ln(\mathrm{Tr}\hat{A})}{\partial \lambda} \left( \hat{A}^2 - \frac{\alpha}{2} \hat{I} \right)$$
$$+ \frac{1}{\mathrm{Tr}\hat{A}} \left( \frac{\partial \hat{A}}{\partial \lambda} \hat{A} + \hat{A} \frac{\partial \hat{A}}{\partial \lambda} - \frac{1}{2} \frac{\partial \alpha}{\partial \lambda} \hat{I} \right). \qquad (16)$$

If $\alpha = 0$, Eq. (16) reduces to

$$\frac{\partial \hat{A}}{\partial \lambda} = -\frac{1}{\mathrm{Tr}\hat{A}} \frac{\partial \ln(\mathrm{Tr}\hat{A})}{\partial \lambda} \hat{A}^2 + \frac{1}{\mathrm{Tr}\hat{A}} \left( \frac{\partial \hat{A}}{\partial \lambda} \hat{A} + \hat{A} \frac{\partial \hat{A}}{\partial \lambda} \right)$$

and by inspection, Eq. (13) holds. On the other hand, if $\alpha \neq 0$ then Eq. (15) can be inverted to give $\hat{I} = 2[\hat{A}^2 - \mathrm{Tr}(\hat{A})\hat{A}]/\alpha$ and Eq. (16) becomes

$$\frac{\partial \hat{A}}{\partial \lambda} = -\frac{\partial \ln(\mathrm{Tr}\hat{A})}{\partial \lambda} \hat{A}$$
$$+ \frac{1}{\mathrm{Tr}\hat{A}} \left( \frac{\partial \hat{A}}{\partial \lambda} \hat{A} + \hat{A} \frac{\partial \hat{A}}{\partial \lambda} - \frac{1}{2} \frac{\partial \alpha}{\partial \lambda} \hat{I} \right)$$
$$- \frac{1}{\mathrm{Tr}\hat{A}} \frac{\partial \ln \alpha}{\partial \lambda} [\hat{A}^2 - \mathrm{Tr}(\hat{A})\hat{A}]. \qquad (17)$$

Inspection of Eq. (17) shows that the SLD is given by Eq. (14). ∎

Some situations require SLDs for operators that act on vector spaces of dimension greater than 2. If such an operator can be decomposed into the sum of operators acting on orthogonal spaces then the quantum Fisher information can be determined in a piecewise fashion.

*Proposition 2.* Suppose that $\hat{\rho} = \sum_k \hat{\rho}_k$ where the supports of $\hat{\rho}_j$ and $\hat{\rho}_k$ are orthogonal whenever $j \neq k$. Let $\hat{L}_k$ be the SLD for $\hat{\rho}_k$, i.e., $\frac{\partial \hat{\rho}_k}{\partial \lambda} = (\hat{L}_k \hat{\rho}_k + \hat{\rho}_k \hat{L}_k)/2$. Then the SLD for

$\hat{\rho}$ is $\hat{L} = \sum_k \hat{L}_k$ and the quantum Fisher information is $H = \sum_k H_k$ where $H_k = \mathrm{Tr}(\frac{\partial \hat{\rho}_k}{\partial \lambda} \hat{L}_k)$.

*Proof.* Suppose that $\frac{\partial \hat{\rho}_k}{\partial \lambda} = (\hat{L}_k \hat{\rho}_k + \hat{\rho}_k \hat{L}_k)/2$. Then

$$\frac{\partial \hat{\rho}}{\partial \lambda} = \sum_k \frac{\partial \hat{\rho}_k}{\partial \lambda} = \sum_k \frac{1}{2}(\hat{L}_k \hat{\rho}_k + \hat{\rho}_k \hat{L}_k). \qquad (18)$$

The support of $\hat{L}_k$ is either the same as or a subspace of the support of $\hat{\rho}_k$. Thus if $j \neq k$ then $\hat{L}_j \hat{\rho}_k = 0$ and

$$\frac{\partial \hat{\rho}}{\partial \lambda} = \frac{1}{2} \left( \sum_j \hat{L}_j \sum_k \hat{\rho}_k + \sum_k \hat{\rho}_k \sum_j \hat{L}_j \right) = \frac{1}{2}(\hat{L}\hat{\rho} + \hat{\rho}\hat{L}). \qquad (19)$$

The result regarding the quantum Fisher information follows by applying similar reasoning to Eq. (14). ∎

### C. Bounds on the quantum Fisher information for phase-flip parameter estimation

A useful tool for addressing bounds on the quantum Fisher information for the phase-flip operation, already considered elsewhere [7,19,21], involves a representation of the channel using an additional ancilla qubit as illustrated in Fig. 2.

This pair of qubits is subjected to the indicated *unitary* operations where the additional "*coin-toss*" unitary operation is

$$\hat{V}(\lambda) := \begin{pmatrix} \sqrt{\lambda} & -\sqrt{1-\lambda} \\ \sqrt{1-\lambda} & \sqrt{\lambda} \end{pmatrix}. \qquad (20)$$

Following the controlled-$Z$ unitary, the ancilla qubit is discarded. Mathematically, this amounts to taking the partial trace over the state space for the ancilla qubit. It is straightforward to show that the resulting transformation of the channel qubit is identical to the phase-flip operation.

The quantum Fisher information of the phase-flip channel alone is bounded from above by the quantum Fisher information of the entire extended channel of Fig. 2 since the set of all possible measurements on both qubits includes the set of all measurements made on the channel qubit alone. When evaluating the quantum Fisher information of the extended channel the controlled-$Z$ unitary and the initial state of the channel qubit can be collectively absorbed into a general measurement procedure since they are parameter-independent. The only relevant entities are the coin-toss unitary and the initial state of the ancilla qubit. Thus the quantum Fisher
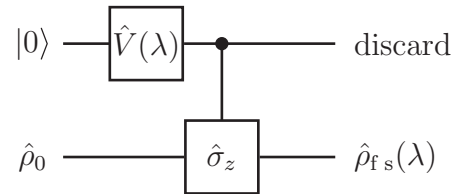


FIG. 2. Unitary representation of a phase-flip channel. The lower qubit is the channel qubit and the upper an ancilla qubit, which is required to be in the initial state, $|0\rangle$. After evolution under the two indicated unitary operations, the ancilla qubit is ignored or discarded.

information for the extended channel is exactly that for the coin-toss unitary applied to the ancilla in the state $|0\rangle$. Proposition 1 and Eq. (12) apply here and give that the quantum Fisher information for the extended channel is $1/\lambda(1-\lambda)$.

Multiple uses of the phase-flip operation can be represented by a similar extension involving multiple coin-toss unitaries and ancilla qubits, each in the initial state $|0\rangle$. According to the scheme above, these parts of the procedure occur independently of each other and it is straightforward to show that the quantum Fisher information of the extended channel (and thus the upper bound on any estimation procedure using any channel input states) using $m$ phase-flip operations is $m/\lambda(1-\lambda)$. This result is independent of the state of the channel qubits and also of any unitaries $\hat{T}_1,\ldots,\hat{T}_m$ that are applied to the channel qubits between channel invocations. Thus *for any parameter estimation scheme involving $m$ channel invocations the quantum Fisher information satisfies*

$$H(\lambda) \leqslant \frac{m}{\lambda(1-\lambda)}. \tag{21}$$

This absolute upper bound has appeared elsewhere in a more general context [19].

### III. INPUT STATE CHOICES AND CONSTRAINTS

The quantum Fisher information partly depends on the choice of the initial state for the system. It will be essential to provide a general mathematical description of the various initial states; the quantum Fisher information and possible constraints on initial states will be expressed partly in terms of parameters appearing in this description.

### A. General product input states

In general the initial state of the system can be assumed to be a product of states for individual qubits; initial entangled states can then be accommodated via appropriate choice of $\hat{T}_0$. The general density operator for an individual qubit has form

$$\hat{\rho} := \tfrac{1}{2}(\hat{I} + \mathbf{r} \cdot \hat{\boldsymbol{\sigma}}) \tag{22}$$

where $\mathbf{r} \cdot \hat{\boldsymbol{\sigma}} = r_x \hat{\sigma}_x + r_y \hat{\sigma}_y + r_z \hat{\sigma}_z$ and $\mathbf{r} = (r_x, r_y, r_z)$ is a conventional three-dimensional real vector whose norm, or *polarization*, satisfies $r \equiv |\mathbf{r}| := \sqrt{r_x^2 + r_y^2 + r_z^2} \leqslant 1$. We assume that all qubits have the same initial polarization. The only remaining possible differences between initial states arises from the orientations of $\mathbf{r}$ for each qubit. By including appropriate single qubit rotations in $\hat{T}_0$, any differences in these orientations can be removed. Thus we assume that $\mathbf{r}$ is identical for each qubit and therefore $\hat{\rho}_{\rm i} = \hat{\rho}_0 \otimes \hat{\rho}_0 \otimes \cdots \otimes \hat{\rho}_0$ ($n$ factors) where

$$\hat{\rho}_0 := \tfrac{1}{2}(\hat{I} + \mathbf{r} \cdot \hat{\boldsymbol{\sigma}}) \tag{23}$$

for a particular given $\mathbf{r}$. The freedom to include single qubit rotations in $\hat{T}_0$ thus implies that the only relevant input state parameter for the quantum Fisher information is the polarization, $r$.

### B. Pure input states

A subset of possible input states are those which are pure, i.e., $r = 1$. Consider a single application of the channel to a qubit in the initial state $|+\hat{\mathbf{y}}\rangle = (|0\rangle + i|1\rangle)/\sqrt{2}$ with all $\hat{T}_j = \hat{I}$. Straightforward application of Proposition 1 and Eq. (12) yields a single use pure state quantum Fisher information

$$H_{\rm s\ pure\ opt}(\lambda) = \frac{1}{\lambda(1-\lambda)}, \tag{24}$$

identical to that of the bound of Eq. (21) with $m = 1$ and therefore optimal. In fact, the projective measurement in the basis $\{|+\hat{\mathbf{y}}\rangle, |-\hat{\mathbf{y}}\rangle\}$, where $|\pm\hat{\mathbf{y}}\rangle = (|0\rangle \pm i|1\rangle)/\sqrt{2}$, is readily shown to yield a classical Fisher information equal to $H_{\rm s\ pure\ opt}(\lambda)$.

It follows that if this is repeated independently with $m$ channel invocations, each applied to a single qubit as described above, the resulting classical Fisher information is $m$ times that for a single channel use. But this saturates the bound of Eq. (21) and thus the optimal pure state quantum Fisher information with $m$ channel invocations is

$$H_{\rm pure\ opt}(\lambda) = \frac{m}{\lambda(1-\lambda)}. \tag{25}$$

Clearly for pure input states the optimal quantum Fisher information can be attained using product states (this is an example of the independent channel use protocol). *For pure input states no correlated state protocol can yield a better estimate than the particular independent channel use protocol describe in this section.* Evidently entanglement or quantum correlations between qubits cannot yield any advantages for phase-flip parameter estimation with initial pure states.

### C. Independent channel use protocol

In the more general independent channel use protocol the initial states are not pure. Suppose that the operation is applied once to each of $m$ qubits, each of which is in the identical initial state, given by Eq. (23) and which appear leftmost in the mathematical representation of the entire system state. The final state of the collection of qubits is $\hat{\rho}_{\rm f} = \hat{\rho}_{\rm f\ s} \otimes \cdots \otimes \hat{\rho}_{\rm f\ s} \otimes \hat{\rho}_0 \otimes \cdots \otimes \hat{\rho}_0$ ($m$ factors of $\hat{\rho}_{\rm f\ s}$) where the final state for the individual qubits to which the operation is applied is $\hat{\rho}_{\rm f\ s} = (1-\lambda)\hat{\rho}_0 + \lambda\hat{\sigma}_z\hat{\rho}_0\hat{\sigma}_z$. Let $\hat{L}_{\rm s}$ be the SLD corresponding to $\hat{\rho}_{\rm f\ s}$ and let $H_{\rm s}(\lambda)$ be the resulting quantum Fisher information for this single qubit. It is straightforward to show that the SLD for all $n$ qubits is $\hat{L}_{\rm s} \otimes \hat{I} \otimes \cdots \otimes \hat{I} + \hat{I} \otimes \hat{L}_{\rm s} \otimes \hat{I} \otimes \cdots \otimes \hat{I} + \cdots$ with one term for each qubit to which the operation is applied. Then Eq. (12) implies that the quantum Fisher information for the entire system is

$$H(\lambda) = m\ H_{\rm s}(\lambda). \tag{26}$$

The quantum Fisher information for a single channel use on one qubit can be computed by first determining the SLD. The conditions of Proposition 1 apply and, since $\hat{\rho}_{\rm f\ s}$ is not a pure state, $\alpha = {\rm Tr}(\hat{\rho}_{\rm f\ s}^2) - ({\rm Tr}\hat{\rho}_{\rm f\ s})^2 \neq 0$. Thus

$$\hat{L}_{\rm s} = 2\left\{\hat{\rho}_{\rm f\ s} - \frac{{\rm Tr}[\hat{\rho}_{\rm f\ s}\hat{\rho}_{\rm f\ s}]}{1 - {\rm Tr}[\hat{\rho}_{\rm f\ s}^2]}(\hat{I} - \hat{\rho}_{\rm f\ s})\right\}. \tag{27}$$

Substituting into Eq. (12) gives

$$H_s(\lambda) = 2\text{Tr}[\hat{\rho}_{\text{f s}}^2] - 2\frac{(\text{Tr}[\hat{\rho}_{\text{f s}}\hat{\rho}_{\text{s}}])^2}{1 - \text{Tr}[\hat{\rho}_{\text{f s}}^2]}. \quad (28)$$

Applying this to $\hat{\rho}_{\text{f s}}$ as produced by the phase-flip channel and using $\hat{\rho}_0$ given by Eq. (23) yields

$$H_s(\lambda) = \frac{4\left(1 - r_z^2\right)\left(r^2 - r_z^2\right)}{(1 - 2\lambda)^2(1 - r^2) + 4\lambda(1 - \lambda)\left(1 - r_z^2\right)}. \quad (29)$$

Thus the optimal single qubit, single channel use quantum Fisher information is attained when $r_z = 0$ and is

$$H_{\text{sopt}}(\lambda) = \frac{4r^2}{1 - (1 - 2\lambda)^2 r^2}. \quad (30)$$

Thus, *for m independent channel uses* the optimal quantum Fisher information is

$$H_{\text{ind opt}}(\lambda) = \frac{4r^2 m}{1 - (1 - 2\lambda)^2 r^2}. \quad (31)$$

One physical implementation which yields a classical Fisher information equal to that of $H_{\text{s opt}}(\lambda)$ commences with $\hat{\rho}_0 = (\hat{I} + r\hat{\sigma}_y)/2$ and follows a single channel invocation with a projective measurement in the basis $\{|+\hat{y}\rangle, |-\hat{y}\rangle\}$ where $|\pm\hat{y}\rangle = (|0\rangle \pm i|1\rangle)/\sqrt{2}$. A straightforward derivation using the probability distribution for the measurement outcomes and Eq. (7) results in $F(\lambda) = H_{\text{s opt}}(\lambda)$.

Equation (31) gives the optimal quantum Fisher information in the independent channel use protocol when all qubits have initial polarization $r$. This provides the benchmark against which correlated state protocol estimation schemes will be compared.

## IV. CORRELATED STATE PROTOCOL PARAMETER ESTIMATION

The results of Sec. II C show that for pure input states, any correlated state protocol procedure cannot improve on the independent channel use protocol quantum Fisher information of Eq. (25). However, for mixed input states, the result of Eq. (31) does not saturate the absolute bound of Eq. (21). Here it is conceivable that a correlated state protocol procedure could improve upon the quantum Fisher information of Eq. (31) given the same number of channel invocations.

Our aim here is to demonstrate that there exists at least one correlated state protocol that provides an enhanced quantum Fisher information. The discussion will focus on one particular protocol and the primary motivation behind the choice of this is that it will be shown to yield gains in the quantum Fisher information over the independent channel use protocol. While this will be sufficient to attain our aims, it is not inconceivable that other such correlated state protocols may exist and that some of these could be more efficient in terms of resource use; these issues are beyond the scope of this article.

Henceforth suppose that $r < 1$. The particular correlated state protocol procedure that we will consider in the remainder of this article is illustrated in Fig. 3. Specifically suppose that $\hat{\rho}_0 = (\hat{I} + r\hat{\sigma}_y)/2$ and $\hat{U}_{\text{prep}}$ is as illustrated in Fig. 4. Denote the density operator for the system after the preparatory gate
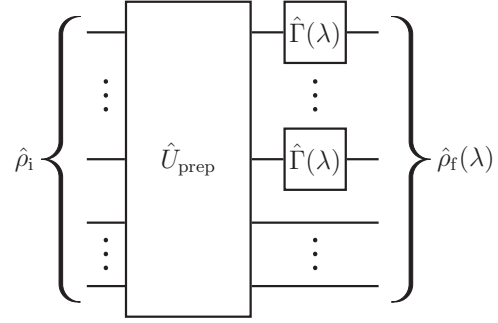


FIG. 3. Single preparation unitary scheme for improved parameter estimation. There are $n$ single qubits. A joint preparatory unitary, $\hat{U}_{\text{prep}}$ precedes a phase-flip operation applied once to each of the uppermost $m$ qubits. This is of the form of the general scheme of Fig. 1 where $\hat{T}_0 = \hat{U}_{\text{prep}}, \hat{T}_m = \hat{I}$ and $\hat{T}_1, \ldots, \hat{T}_{m-1}$ being swap gates between pairs of qubits.

by $\hat{\rho}_{\text{prep}} := \hat{U}_{\text{prep}}\hat{\rho}_i\hat{U}_{\text{prep}}^\dagger$. Then (see Appendix A)

$$\hat{\rho}_{\text{prep}} = \sum_{x=0}^{(N-1)/2} \hat{\rho}_{(x)} \quad (32)$$

where $N = 2^n - 1$ and

$$\hat{\rho}_{(x)} = \left[\frac{f(x) + f(N - x)}{2}\right] [|x\rangle\langle x| + |N - x\rangle\langle N - x|]$$
$$+ i\left[\frac{f(x) - f(N - x)}{2}\right] [|x\rangle\langle N - x| - |N - x\rangle\langle x|]. \quad (33)$$

In Eq. (33), $|x\rangle = |x_n \ldots x_1\rangle$ where $x_n \ldots x_1$ is the binary representation of $x$, with the rightmost bit the least significant digit. Note that $|N - x\rangle$ can be obtained by flipping each bit value in $|x\rangle$. Finally

$$f(x) = \frac{(1 + r)^j (1 - r)^{n-j}}{2^n} \quad (34)$$

where $j$ is the number of zeros in the bit string for $x$. The support of $\hat{\rho}_{(x)}$ is the two-dimensional subspace spanned by
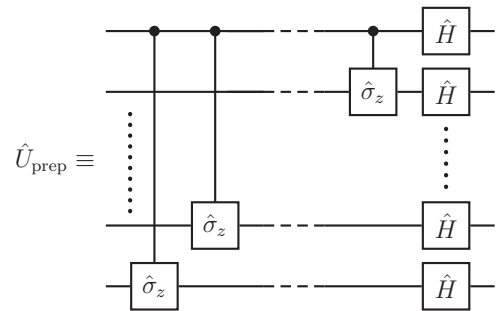


FIG. 4. The preparatory gate, $\hat{U}_{\text{prep}}$, consists of a collection of controlled-$Z$ operations applied to each distinct pair of qubits followed by a Hadamard transformation applied to each qubit. The preparatory gate is symmetric under interchange of any qubits since a controlled-$Z$ gate is symmetric between the pair of qubits on which it acts.

$\{|x\rangle, |N - x\rangle\}$ and, with one type of exception, if $x \neq x'$ then $\hat{\rho}_{(x)}$ and $\hat{\rho}_{(x')}$ have orthogonal supports. The only exceptions to this are that $\hat{\rho}_{(x)} = \hat{\rho}_{(N-x)}$, a fact which is straightforward to show. Mathematically, the preparatory gate has engineered a splitting of the Hilbert space for the entire system into mutually orthogonal two-dimensional subspaces.

The preparation scheme is symmetric under interchange of qubits and, without loss of generality, consider the situation where the phase-flip operation is applied once to each of the qubits corresponding to the $m$ least significant (or rightmost) digits in the computational basis representation. Then

$$\hat{\rho}_{\mathrm{f}} = \sum_{x=0}^{(N-1)/2} \hat{\rho}_{\mathrm{f}\,(x)} \qquad (35)$$

where $\hat{\rho}_{\mathrm{f}\,(x)}$ results from applying the phase-flip operation to $\hat{\rho}_{(x)}$. Here

$$
\begin{aligned}
|x\rangle\langle x| &\mapsto |x\rangle\langle x| \\
|N - x\rangle\langle N - x| &\mapsto |N - x\rangle\langle N - x| \\
|x\rangle\langle N - x| &\mapsto (1 - 2\lambda)^m |x\rangle\langle N - x| \\
|N - x\rangle\langle x| &\mapsto (1 - 2\lambda)^m |N - x\rangle\langle x|
\end{aligned}
\qquad (36)
$$

and thus

$$
\begin{aligned}
\hat{\rho}_{\mathrm{f}\,(x)} =& \left[\frac{f(x) + f(N - x)}{2}\right] \left[|x\rangle\langle x| + |N - x\rangle\langle N - x|\right] \\
&+ i(1 - 2\lambda)^m \left[\frac{f(x) - f(N - x)}{2}\right] \\
&\times \left[|x\rangle\langle N - x| - |N - x\rangle\langle x|\right].
\end{aligned}
\qquad (37)
$$

Each of $\hat{\rho}_{\mathrm{f}\,(x)}$ has a two-dimensional support and these are clearly orthogonal for distinct values of $x$ in the range $0, \ldots, (N - 1)/2$. Thus Proposition 2 gives that $H(\lambda) = \sum_{x=0}^{(N-1)/2} H_{(x)}$ where $H_{(x)} = \mathrm{Tr}(\frac{\partial \hat{\rho}_{\mathrm{f}(x)}}{\partial \lambda} \hat{L}_{(x)})$ and $\hat{L}_{(x)}$ is the score operator for $\hat{\rho}_{\mathrm{f}(x)}$.

Proposition 1 applies to $\hat{\rho}_{\mathrm{f}(x)}$ since it has two-dimensional support and $\mathrm{Tr}\hat{\rho}_{\mathrm{f}(x)} = f(x) + f(N - x) \neq 0$. Direct calculation shows that $\mathrm{Tr}(\hat{\rho}_{\mathrm{f}(x)})^2 - (\mathrm{Tr}\hat{\rho}_{\mathrm{f}(x)})^2 = 0$ if and only if $[f(x) - f(N - x)]^2 (1 - 2\lambda)^{2m} = [f(x) + f(N - x)]^2$. The fact that $f(x) \geqslant 0$ implies that $[f(x) + f(N - x)]^2 \geqslant [f(x) - f(N - x)]^2$. Furthermore $(1 - 2\lambda)^{2m} \leqslant 1$. This implies $\mathrm{Tr}(\hat{\rho}_{\mathrm{f}(x)})^2 - (\mathrm{Tr}\hat{\rho}_{\mathrm{f}(x)})^2 = 0$ if and only if $(1 - 2\lambda) = \pm 1$ (i.e., $\lambda = 0, 1$) and $[f(x) + f(N - x)]^2 = [f(x) - f(N - x)]^2$ [i.e., $f(x) = 0$ or $f(N - x) = 0$]. The latter is only true when $r = 1$. By the assumption that $r < 1$, $\mathrm{Tr}(\hat{\rho}_{\mathrm{f}(x)})^2 - (\mathrm{Tr}\hat{\rho}_{\mathrm{f}(x)})^2 \neq 0$ and Proposition 1 gives that $\hat{L}_{(x)}$ derives from Eq. (14). This together with Eq. (12) yields

$$H(\lambda) = 4m^2(1 - 2\lambda)^{2m-2} \sum_{x=0}^{(N-1)/2} \frac{[f(x) - f(N - x)]^2\,[f(x) + f(N - x)]}{[f(x) + f(N - x)]^2 - (1 - 2\lambda)^{2m}\,[f(x) - f(N - x)]^2}. \qquad (38)$$

The term within the summation of Eq. (38) yields the same quantity when $x$ is replaced by $N - x$. Thus the range of the sum can be extended to 0 to $N$ provided that the result is halved. Also the term within the summation only depends on the number of zeros in the bit string for $x$ and different $x$ can give the same contribution. Let $j$ denote a possible number of zeros. For $x$ ranging from 0 to $N$ the number of terms whose bit strings have $j$ zeros is $\binom{n}{j}$. For each of these terms $f(x) \pm f(N - x) = [(1 + r)^j(1 - r)^{n-j} \pm (1 + r)^{n-j}(1 - r)^j]/2^n$. Thus

$$H(\lambda) = \frac{m^2(1 - 2\lambda)^{2m-2}}{2^{n-1}} \sum_{j=0}^{n} \binom{n}{j} \frac{c_j^2 d_j}{d_j^2 - (1 - 2\lambda)^{2m} c_j^2}. \qquad (39)$$

where

$$c_j = (1 + r)^j(1 - r)^{n-j} - (1 + r)^{n-j}(1 - r)^j, \qquad (40a)$$

$$d_j = (1 + r)^j(1 - r)^{n-j} + (1 + r)^{n-j}(1 - r)^j. \qquad (40b)$$

Equations (39) and (40) form the central result from which the principal conclusions of this article follow.

The estimation accuracies of this correlated state protocol and the optimal independent channel use protocol can be compared using the quantum Fisher informations of Eqs. (39) and (31). The constraints are that *the polarizations, channel parameter values and number of channel invocations are identical for each protocol, respectively*. Given this, the

fractional gain provided by the correlated state protocol is

$$G(\lambda) := \frac{H(\lambda)}{H_{\mathrm{ind\ opt}}(\lambda)} \qquad (41)$$

and, when $G(\lambda) > 1$, the correlated state protocol performs more accurately than the independent channel use protocol. Equations (31) and (39) give

$$G(\lambda) = \frac{m(1 - 2\lambda)^{2m-2}}{2^{n+1}r^2} \sum_{j=0}^{n} \binom{n}{j} \frac{c_j^2}{d_j} \frac{1 - (1 - 2\lambda)^2 r^2}{1 - (1 - 2\lambda)^{2m} c_j^2/d_j^2}. \qquad (42)$$

For a fixed polarization $0 \leqslant r < 1$ the gain is a monotonically increasing function of $(1 - 2\lambda)^2$. This follows from the fact that the gain function is a sum of terms of the form $(1/x - r^2)/(1/x^m - c_j^2/d_j^2)$ with $x = (1 - 2\lambda)^2$. Then, as shown in Appendix B, $c_j^2/d_j^2 \geqslant r^2$ (the only exception occurs when $j = n/2$, in which case $c_j = 0$ and this term is irrelevant in the sum). Thus as $x$ increases from 0 to 1, the fraction $(1/x - r^2)/(1/x^m - c_j^2/d_j^2)$ increases. It follows that *for a fixed polarization the minimum gain is*

$$G_{\min} = \begin{cases} \frac{1}{2^{n+1}r^2} \sum_{j=0}^{n} \binom{n}{j} \frac{c_j^2}{d_j} & \text{if } m = 1 \\ 0 & \text{if } m \neq 1 \end{cases} \qquad (43)$$

and in each case these are attained when $\lambda = 1/2$. Similarly *for a fixed polarization the maximum gain is*

$$G_{\max} = \frac{m}{2^{n+1}r^2} \sum_{j=0}^{n} \binom{n}{j} \frac{c_j^2}{d_j} \frac{1-r^2}{1-c_j^2/d_j}. \qquad (44)$$

and this is attained when $\lambda = 0$ or $1$.

The other extremes of the parameter space are $r \to 0$ and $r \to 1$. In the first case, equivalent to $r \ll 1, c_j \approx 2r(2j-n)$ and $d_j \approx 2$. Equation (42) then implies that

$$G \to mn(1-2\lambda)^{2m-2} \quad \text{as} \quad r \to 0. \qquad (45)$$

As $r \to 1$, $c_j \to 0$ unless $j = 0$ or $n$. Specifically $c_0 = -c_n \to -2^n$ and $d_0 = d_n \to 2^n$. Thus Eq. (42) implies that

$$G \to m(1-2\lambda)^{2m-2} \frac{1-(1-2\lambda)^2}{1-(1-2\lambda)^{2m}} \quad \text{as} \quad r \to 1. \qquad (46)$$

Note that this excludes the cases where both $\lambda = 0$ or $1$ and $r = 1$ here as Proposition 1 will give an SLD and Fisher information different to that of Eq. (38). These cases are beyond consideration here.

The preceding results indicate important differences between situations in which the phase-flip operation is invoked only once and those where it is invoked more than once.

## V. ESTIMATION WITH A SINGLE CHANNEL USE

If the phase-flip is invoked only once then the correlated state protocol yields the following gain

$$G(\lambda) = \frac{1}{2^{n+1}r^2} \sum_{j=0}^{n} \binom{n}{j} \frac{c_j^2}{d_j} \frac{1-(1-2\lambda)^2 r^2}{1-(1-2\lambda)^2 c_j^2/d_j^2}. \qquad (47)$$

As shown in Appendix B, with just one phase-flip operation and for any $n \geqslant 2$, $G(\lambda) > 1$ if $0 < r < 1$. This shows that, *for single phase-flip invocation, the correlated state protocol of Sec. IV gives a more accurate estimate of the phase-flip parameter than the independent channel use protocol whenever the polarizations used by each are identical*. Additionally Eqs. (46) and (45) imply that $G \to 1$ as $r \to 1$ and that $G \to n$ as $r \to 0$. The striking observation is that *for sufficiently small polarizations and a single phase-flip invocation, the correlated state protocol increases the estimation accuracy by a factor of approximately n regardless of the value of the parameter.*

For polarization values that do not approach these limits, a relevant measure of success would be the minimum gain, which is, according to Eq. (43),

$$G_{\min} = \frac{1}{2^{n+1}r^2} \sum_{j=0}^{n} \binom{n}{j} \frac{c_j^2}{d_j} \qquad (48)$$

and is larger than 1.

Typical plots of the gain, which illustrate the features described above, are provided in Fig. 5. These indicate that for a fixed value of $\lambda$, the gain does not necessarily increase or decrease with increasing value of $r$, although it appears that $G_{\min}$ decreases as $r$ increases. Also, depending on the polarization and parameter value, it is possible to attain gains in excess of $n$. For example, with $n = 2$ the extremes are $G_{\min} = 2/(1+r^2)$ and $G_{\max} = 2(1+r^2)/(1-r^2)$. Clearly a
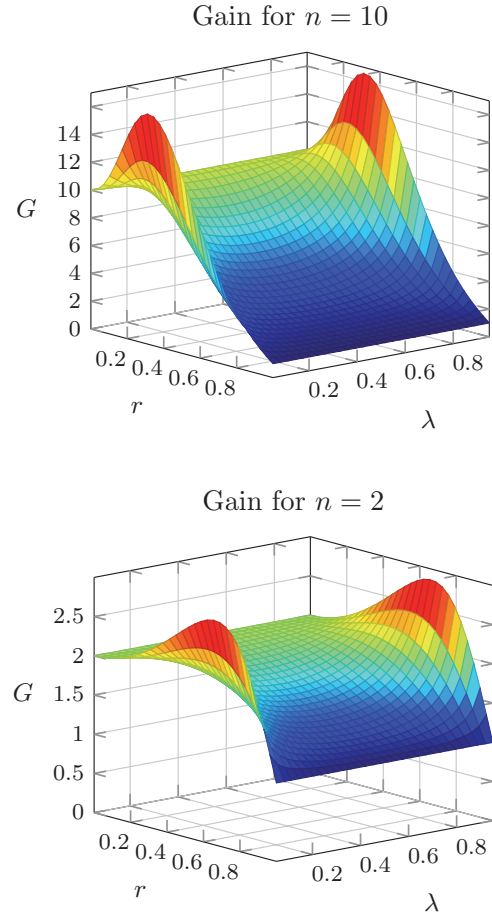


FIG. 5. (Color online) Gain using one phase-flip invocation. Both plots are over the range $0.05 \leqslant \lambda \leqslant 0.95$ and $0 \leqslant r \leqslant 1$.

gain of arbitrarily large magnitude can be attained provided that, for example, $\lambda$ and $r$ are sufficiently close to 1.

## VI. ESTIMATION WITH MULTIPLE CHANNEL USES

In the more general case of the protocol of Figs. 3 and 4, there are $m$ phase-flip invocations on $n$ qubits. Here Eq. (42) indicates that $G(1/2) = 0$ and thus there will be ranges of phase-flip parameters for which the correlated state protocol provides a gain of less than 1 and thus gives a lower accuracy than the independent channel use protocol.

A relatively simple example has $n = 2$ and $m = 2$ and the gain is illustrated in Fig 6. This indicates regions in the parameter space for which $G(\lambda) > 1$ and other regions for which $G(\lambda) < 1$. Comparing Figs. 5 and 6 for $n = 2$ indicate regions for values of $\lambda$ near to 0 or 1 in which the gain using two phase-flip invocations is larger than that using one phase-flip invocation.

Whenever $m \geqslant 2$, Eq. (43) precludes any advantages for the correlated state protocol over the *entire range* of $\lambda$. Similarly Eq. (45) results in the same loss of advantage over the entire parameter range when $r \to 0$. However, Eq. (44) shows that as $\lambda \to 0$ or $\lambda \to 1$ the gain approaches $m$ times that of the correlated state protocol which uses only one phase-flip invocation. Thus for such parameter values the correlated
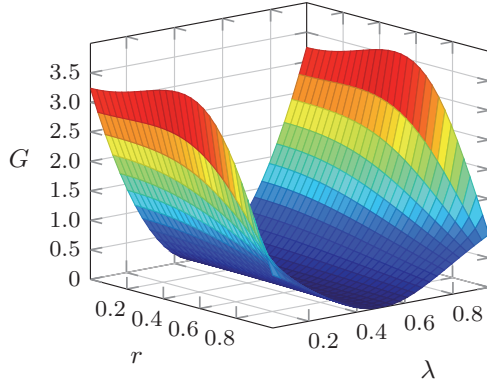
FIG. 6. (Color online) Gain using two phase-flip invocations on two qubits, plotted for $0.05 \leqslant \lambda \leqslant 0.95$ and $0 \leqslant r \leqslant 1$.

state protocol will be superior to the independent channel use protocol. These facts are illustrated in Fig. 7.

It is possible to attain a substantial gain. For example, if $r \ll 1$, then Eq. (45) shows that $G(\lambda) \geqslant n$ provided that

$$\lambda \leqslant \frac{1}{2} \left[ 1 - \frac{1}{m^{1/(2m-2)}} \right]. \tag{49}$$

The right hand side of this decreases as $m$ increases. Note that a symmetrical situation occurs by replacing $\lambda$ by $1 - \lambda$.

Of special interest are the cases where the phase-flip is applied to all qubits (i.e., $m = n$). Here, the resulting upper bounds on $\lambda$ such that $G(\lambda) \geqslant n$ are illustrated in Fig. 8. This situation appears in NMR implementations of quantum information processing, where all of the qubits are subjected to the same or similar phase-flip operations. In room-temperature solution-state NMR implementations [14], the polarization of a single qubit is typically $r \approx 10^{-4}$, which satisfies the assumptions resulting in Fig. 8. Considering, $\lambda = 0.10$, a typical value appearing in Fig. 8, Eq. (3) shows that $\lambda \leqslant 0.10$ when $t \leqslant 0.22T_2$. In typical NMR quantum information processing experiments [31], $T_2 \approx 1$ s and thus by limiting the period of evolution of the system to less than about 0.2 s, the gain attained in estimating $\lambda$ will exceed $n$. This is certainly feasible to attain.

## VII. ENTANGLEMENT AND DISCORD CONSIDERATIONS

It is customary to inquire about specifically which quantum resources could be responsible for gains of the type provided by the correlated state protocol. In other instances where quantum systems offer gains for parameter estimation, attention has focused on the presence of entanglement or other types of quantum correlations [8,26]. In the context of mixed state phase estimation it was shown that entanglement was not necessary for improved accuracy but it appears that quantum discord, a more general form of quantum correlation, is present whenever there is a gain in accuracy [26].

The preparation scheme for Fig. 4 is capable of producing states for which the measurement outcomes are correlated and it is of interest to determine the extent to which various quantum correlations are present in our scheme. The simplest context in which to address this for the correlated state protocol
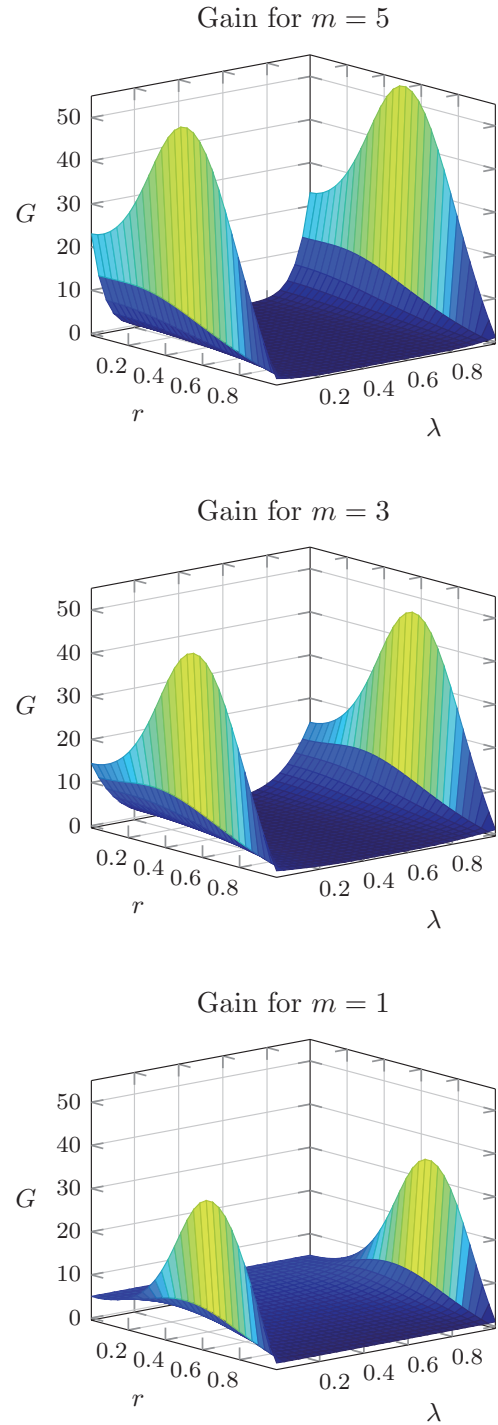


FIG. 7. (Color online) Gain using various numbers of phase-flip invocations on five qubits, plotted for $0.0005 \leqslant \lambda \leqslant 0.9995$ and $0 \leqslant r \leqslant 1$.

is that involving two qubits. For $n = 2$ the gain is

$$G(\lambda) = 2m(1 - 2\lambda)^{2m-2}(1 + r^2) \frac{1 - (1 - 2\lambda)^2 r^2}{(1 + r^2)^2 - 4r^2(1 - 2\lambda)^{2m}}. \tag{50}$$

We aim to determine whether the presence of entanglement is necessary for $G(\lambda) > 1$ and whether other types of quantum correlation may be responsible for the gain attained by the
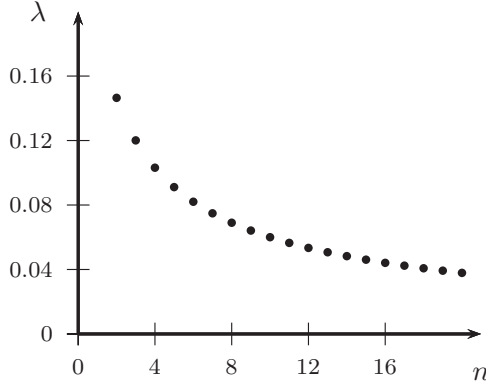
FIG. 8. Maximum value of $\lambda$ vs $n$ such that $G(\lambda) \geqslant n$ for cases where $r \ll 1$.

correlated state protocol. We offer a preliminary investigation of these issues, focusing mostly on the presence of these correlations. A more complete treatment, which may consider issues such as the dependence of discord on the purity of the state and the extent of classical correlations in these protocols, is beyond the scope of this article.

The existence of entanglement or other forms of quantum correlation is inferred from the density operator for the system. After $m$ channel invocations, Eq. (37) yields that, in the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$,

$$\hat{\rho}_f = \frac{1}{2} \begin{pmatrix} \frac{1+r^2}{2} & 0 & 0 & ir\mu \\ 0 & \frac{1-r^2}{2} & 0 & 0 \\ 0 & 0 & \frac{1-r^2}{2} & 0 \\ -ir\mu & 0 & 0 & \frac{1+r^2}{2} \end{pmatrix}. \tag{51}$$

where $\mu := (1 - 2\lambda)^m$. Note that the system density operator prior to channel invocation is the same as that for which $\lambda = 0$, or, equivalently, $\mu = 1$.

### A. Entanglement and separability

A necessary and sufficient condition for a two qubit state to be separable (nonentangled) is that the partial transpose of the density operator for the state is a positive operator [32,33]. The partial transpose of $\hat{\rho}_f$ is

$$\hat{\rho}_f^{PT} = \frac{1}{2} \begin{pmatrix} \frac{1+r^2}{2} & 0 & 0 & 0 \\ 0 & \frac{1-r^2}{2} & ir\mu & 0 \\ 0 & -ir\mu & \frac{1-r^2}{2} & 0 \\ 0 & 0 & 0 & \frac{1+r^2}{2} \end{pmatrix}. \tag{52}$$

This will be a positive operator if and only if its eigenvalues are all positive. The block diagonal form of this matrix renders the computation of the eigenvalues straightforward and it can be shown that these are positive if and only if $0 < r < \sqrt{(1-2\lambda)^{2m} + 1} - |(1-2\lambda)^m|$. In particular, the separability of the state prior to channel invocation can be checked by setting $\lambda = 0$ in this condition. This reveals that the input state prior to channel invocation is separable if and only if $0 < r < \sqrt{2} - 1$ and the channel operations cannot change this since they are local. However, for such polarizations, the results of Secs. V and VI indicate the correlated state protocol

can yield greater accuracy than the independent channel use when $m = 1$. Thus *the gains of the correlated state protocol cannot be ascribed to entanglement.*

### B. Discord

Quantum discord is an alternative and more general characterization of the correlations between two quantum systems [34,35]. Discord requires the density operator for the bipartite system, $\hat{\rho}$ and the reduced density operators for the two the subsystems $\hat{\rho}^A$ and $\hat{\rho}^B$, each of which is attained by taking the partial trace over the complementary system. The quantum mutual information is $\mathcal{I}(\hat{\rho}) := S(\hat{\rho}^A) + S(\hat{\rho}^B) - S(\hat{\rho})$ where the von Neumann entropy is $S(\hat{\rho}) := -\text{Tr}(\hat{\rho} \log_2 \hat{\rho})$.

A possible alternative approach to defining the quantum mutual information supposes that a measurement, described by POVM operators $\{\hat{E}_k\}$ where $k$ indexes the possible measurement outcomes, has been performed on system B and then inquires about the remaining information in subsystem A. Outcome $k$ occurs with probability $p_k := \text{Tr}(\hat{E}_k \hat{\rho})$ and the state of system A following this outcome is $\hat{\rho}_{A|k} := \text{Tr}_B(\hat{\rho}\hat{E}_k)/p_k$. If the systems were in a product state prior to this measurement, then $\hat{\rho}^A = \hat{\rho}_{A|k}$ for all $k$. Thus a measure of the classical correlation between the subsystems, conditional on this particular measurement, can be quantified by $S(\hat{\rho}^A) - \sum_k p_k S(\hat{\rho}_{A|k})$, where the sum is an average over all possible measurement outcomes. Let $S(\hat{\rho}|\{\hat{E}_k\}) := \sum_k p_k S(\hat{\rho}_{A|k})$. The overall classical correlation between the subsystems is them an maximum over all possible measurements and is $\mathcal{C}(\hat{\rho}) := S(\hat{\rho}^A) - \min S(\hat{\rho}|\{\hat{E}_k\})$ where the minimization is done over all POVMs on the subsystem B. The quantum discord is defined as the difference between the quantum mutual information and the classical correlation, $\mathcal{Q} := \mathcal{I}(\hat{\rho}) - \mathcal{C}(\hat{\rho})$. In general, $\mathcal{Q}(\hat{\rho}) \geqslant 0$ and it is invariant under local unitary transformations on the subsystems [34,35].

The requirement for minimization over all possible measurements means that computing the quantum discord is generally very difficult. However, analytical expressions exist for certain classes of two-qubit states [36,37]. In particular an expression exists for for "X states," which have the form

$$\hat{\rho} = \frac{1}{4} \left( \hat{I} + \sum_{j=1}^{3} c_j \hat{\sigma}_j \otimes \hat{\sigma}_j \right) \tag{53}$$

where $c_j$ are all real and $\hat{\sigma}_1 = \hat{\sigma}_x, \ldots$. Let

$$\lambda_0 := 1 - c_1 - c_2 - c_3, \tag{54a}$$
$$\lambda_1 := 1 - c_1 + c_2 + c_3, \tag{54b}$$
$$\lambda_2 := 1 + c_1 - c_2 + c_3, \tag{54c}$$
$$\lambda_3 := 1 + c_1 + c_2 - c_3, \tag{54d}$$

and

$$c := \max \{|c_1|, |c_2|, |c_3|\}. \tag{55}$$

Then [36]

$$\mathcal{Q}(\hat{\rho}) = \frac{1}{4} \sum_{j=0}^{3} \lambda_j \log_2 \lambda_j - \frac{1-c}{2} \log_2 (1-c)$$
$$- \frac{1+c}{2} \log_2 (1+c). \tag{56}$$

The final density operator of Eq. (51) can be brought into the form (as it is the relevant values of $c_j$ would not all be real) of Eq. (53) by the single qubit unitary

$$\hat{U} = \begin{pmatrix} 0 & e^{i\pi/8} \\ e^{-i\pi/8} & 0 \end{pmatrix} \tag{57}$$

applied to each qubit. This will not alter the quantum discord. After this, direct calculation gives

$$\lambda_0 := 1 - r^2, \tag{58}$$

$$\lambda_1 := 1 + 2r(1 - 2\lambda)^m + r^2, \tag{59}$$

$$\lambda_2 := 1 - 2r(1 - 2\lambda)^m + r^2, \tag{60}$$

$$\lambda_3 := 1 - r^2, \tag{61}$$

and

$$c := \max\{r^2, r(1 - 2\lambda)^m\}. \tag{62}$$

The discord for $\hat{\rho}_{\text{prep}}$ can be determined by setting $\lambda = 0$ in these expressions. Then $c = r$ and

$$\mathcal{Q}(\hat{\rho}) = \frac{1+r}{2}\log_2(1+r) + \frac{1-r}{2}\log_2(1-r). \tag{63}$$

If $r > 0$, then $\frac{\partial \mathcal{Q}}{\partial r} > 0$ and, since $\mathcal{Q} = 0$ when $r = 0$, this implies that the system state prior to the channel invocation has nonzero quantum discord. In this sense, there are always quantum correlations present prior to channel invocation.

After channel invocation, the discord depends on whether $r > \mu := (1 - 2\lambda)^m$ or not. In all cases, it is clear that replacing $\mu$ with $-\mu$ does not alter $\mathcal{Q}$ and thus it suffices to consider $\mu > 0$ (negative values are possible if $m = 1$). In Appendix C it is shown that for $\mu > 0$

$$\frac{\partial \mathcal{Q}}{\partial \mu} > 0 \quad \text{and} \quad \frac{\partial \mathcal{Q}}{\partial r} > 0. \tag{64}$$

With regard to $\lambda$, one extreme is $\lambda = 1/2$ (i.e., $\mu = 0$). Here $c = r^2$ and

$$\mathcal{Q}(\hat{\rho}) = 0. \tag{65}$$

Thus regardless of the number of channel invocations there is no quantum discord *after channel invocation* when $\lambda = 1/2$ (the final state of the system is clearly separable). However, with a single channel invocation, the gain in quantum Fisher information is always at least 1 and approaches 2 as $r \to 0$. Thus *the gains provided by the correlated state protocol cannot always be attributed to the existence of nonzero quantum discord after the channel invocation and before any measurements but do appear to be associated with nonzero discord before channel invocation*. This is in contrast to the situation in mixed state phase shift estimation in which, whenever there are gains, there is quantum discord present after channel invocation and this equals the quantum discord before channel invocation [26]. In the correlated state protocol, discord has been consumed by invoking the channel. This is evident from Eq. (64), which implies that when $\lambda = 0$ the discord is at a maximum (this corresponds to $\mu = 1$); this is also the discord of the system prior to channel invocation. Then for any nontrivial channel ($\lambda \neq 0,1$) the discord after channel invocation is lower than before. Consumption of discord by nonunitary channels has been observed elsewhere [38,39] and,

TABLE I. Values of $r$ such that $\frac{\partial G}{\partial r} = 0$ as calculated using Eq. (66).

| $m$ | $\lambda$ | $r$ |
|---|---|---|
| 1 | 0.95 | 0.66 |
| 1 | 0.99 | 0.83 |
| 2 | 0.95 | 0.48 |
| 2 | 0.99 | 0.76 |

for these scenarios, bounds relating discord consumption and improved information processing exist. Whether there are such relationships for parameter estimation are beyond the scope of this article but may warrant further investigation.

Beyond this extreme, i.e., $\mu > 0$, the discord is nonzero. For a fixed value of $\lambda \neq 0$, the discord is a monotonically increasing function of $r$. However, Eq. (50) can be used to show that the gain is not monotonically increasing. This requires finding $r$ such that $\frac{\partial G}{\partial r} = 0$. Explicit calculation reveals that this occurs when $r = 0$ or

$$r^2 = \frac{1 + \nu \pm 2\sqrt{(1 + \nu)^2 \nu^m - 4\nu^{m+1}}}{4\nu^{2m+1} - 1 - \nu} \tag{66}$$

where $\nu = (1 - 2\lambda)^2$. Table I lists examples where $r$ is such that $\frac{\partial G}{\partial r} = 0$.

This indicates that the gain is not a monotonically increasing function of $r$. Thus there will be regions where the gain attained decreases while the discord of the post-channel invocation state increases but other regions where the gain increases while this discord increases. This is also evident from the lower graph of Fig. 5.

The landscape with respect to the use of discord for estimation gains with mixed state correlated protocols of the type described here appears to be more contorted for phase-flip estimation than for phase estimation. The nature of these contortions and why they arise is beyond the immediate aims of this article but would be worthy of additional investigation.

### C. Polarization

The quantum Fisher information provided by the independent channel use protocol clearly increases as the polarization increases according to Eq. (31). A possible explanation for gains in this provided by the correlated state protocol could be that this protocol has increased the polarization of each qubit. However, inspection of $\hat{\rho}_{\text{prep}}$ [which is attained by setting $\lambda = 0$ in Eq. (51)] reveals that the polarization of each individual qubit after application of the preparatory unitary is zero. Thus these gains cannot be associated with polarization enhancement attained by the preparatory unitary.

It should be possible to attain gains by applying algorithmic cooling to a single qubit [40,41]. This would increase the qubit polarization, thereby increasing the quantum Fisher information. Such algorithmic cooling procedures involve an open system in which system qubits are repeatedly discarded and replaced with fresh qubits in thermal equilibrium and appear to be very different to the closed system that is used by the correlated state protocol which is the focus of this article. It appears to be unlikely that the gains attained by the correlated state protocol can be attributed to any correspondence

with those that could be attained using algorithmic cooling. However, how two such approaches would compare in terms of resources and performance would be worthy of future investigation.

## VIII. CONCLUSION

We have considered two general phase-flip parameter estimation protocols and compared their accuracies as quantified by the quantum Fisher information. The first protocol, the independent state protocol, proceeds in such a way that the states of the system qubits are always nonentangled. This is akin to the classical strategy of repeating the estimation procedure independently multiple times and offers the same quantitative advantages. The second protocol, uses states which are correlated via an entangling unitary prior to channel invocation. If the number of channel invocations is identical and the polarizations are identical, we have shown that the correlated state protocol can improve on the accuracy of the parameter estimation when compared to the independent state protocol. The extent of the improvement depends on the actual parameter value but it can be substantially larger than the typical gain attained in the classical strategy.

Our analysis demonstrates that the correlated state protocol attains advantages when the state of the system after channel invocation and prior to measurement is separable and, in some cases, where there is zero quantum discord after channel invocation. The gains afforded by the correlated state protocol cannot be attributed entirely to entanglement or nonzero quantum discord after channel invocation and it is unclear which quantum resources are possibly responsible for the gains. The relationship between quantum discord and gains in estimation accuracy appears to be more involved than that for mixed-state phase estimation. This relationship remains to be investigated more completely.

The correlated state protocol may be of use in estimating the dephasing time, $T_2$ of a spin in NMR as this is closely related to the phase-flip parameter, $\lambda$. However, we have assumed that the phase-flip parameter is identical for all qubits; this is not necessarily true in implementations. For example, NMR systems used in quantum information processing provide different dephasing times among the system qubits (for examples from NMR quantum information processing see Refs. [42–46]). It remains to be seen what gains methods like those of this article will yield in such cases.

We stress that the correlated state protocol that we analyzed uses one particular preparatory unitary, $\hat{U}_{\text{prep}}$, and we have not determined whether this is optimal.

Finally, the methods used here are applicable to other Pauli-type channels, i.e., where $\hat{\sigma}_z$ is replaced by $\sigma_n$ with $n$ corresponding to any direction, in Eq. (1). These channels are all related by single qubit rotations and the latter can be absorbed into the various parameter independent unitaries and measurements used in the estimation protocol.

## ACKNOWLEDGMENT

## APPENDIX A: STATE AFTER PREPARATORY UNITARY

The Hadamard gate of Fig. 4 acts on a single qubit as

$$|z_j\rangle \rightarrow \frac{1}{\sqrt{2}} \sum_{y_j=0}^{1} (-1)^{z_j y_j} |y_j\rangle \tag{A1}$$

and the controlled-$Z$ between bits $j$ and $k$ via

$$|z_n \dots z_1\rangle \rightarrow (-1)^{z_j z_k} |z_n \dots z_1\rangle. \tag{A2}$$

Thus

$$\hat{U}_{\text{prep}} = \frac{1}{2^{n/2}} \sum_{y,z} (-1)^{z \cdot y + s} |y\rangle\langle z| \tag{A3}$$

where the binary representation of $z$ is $z_n \dots z_1$ and the sum is over all possible values of these binary representations of $z$ and $y$. Also $z \cdot y := \sum_{j=1}^{n} z_j y_j$ and $s := z_1 z_2 + z_1 z_3 + \cdots + z_{n-1} z_n$, in which each distinct pair of nonequal subscripts appears exactly once.

In order to determine $\hat{\rho}_{\text{prep}} := \hat{U}_{\text{prep}} \hat{\rho}_i \hat{U}_{\text{prep}}^\dagger$, it will be convenient to reexpress $\hat{\rho}_i = \hat{\rho}_0 \otimes \hat{\rho}_0 \otimes \cdots \otimes \hat{\rho}_0$ in terms of the single qubit basis states, $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. Here

$$\hat{\rho}_0 = q|+\rangle\langle+| + (1-q)|-\rangle\langle-| \tag{A4}$$

where $q = (1+r)/2$. A convenient representation of $\hat{\rho}_i$ is in terms of a sum of mutually orthogonal operators, each of which has a two-dimensional support. Thus

$$\hat{\rho}_i = \sum_{x=0}^{(N-1)/2} \hat{\rho}_{i(x)} \tag{A5}$$

where straightforward algebra using Eq. (A4) gives

$$\hat{\rho}_{i(x)} = f(x)|\overline{x}\rangle\langle\overline{x}| + f(N-x)|\overline{N-x}\rangle\langle\overline{N-x}| \tag{A6}$$

with $f$ as defined in Eq. (34). Here $|\overline{x}\rangle$ is such that if $x_n \dots x_1$ is the binary representation of $x$, then $|\overline{x}\rangle$ contains the binary string for $x$ with "0" replaced by "+" and "1" replaced by "−". For example with $n = 3$, $|\overline{0}\rangle = |+++\rangle = |+\rangle|+\rangle|+\rangle, |\overline{1}\rangle = |++-\rangle, |\overline{2}\rangle = |+-+\rangle$, etc. In this example $\rho_{i(0)}$ has support spanned by $\{|+++\rangle, |---\rangle\}$, $\rho_{i(1)}$ has support spanned by $\{|++-\rangle, |--+\rangle\}$ and so on.

The preparatory unitary results in

$$\hat{\rho}_{\text{prep}} = \sum_{k=0}^{(N-1)/2} \hat{\rho}_{(x)} \tag{A7}$$

where

$$\begin{aligned}\hat{\rho}_{(x)} &:= \hat{U}_{\text{prep}} \hat{\rho}_{i(x)} \hat{U}_{\text{prep}}^\dagger \\ &= f(x)\hat{U}_{\text{prep}}|\overline{x}\rangle\langle\overline{x}|\hat{U}_{\text{prep}}^\dagger \\ &\quad + f(N-x)\hat{U}_{\text{prep}}|\overline{N-x}\rangle\langle\overline{N-x}|\hat{U}_{\text{prep}}^\dagger. \end{aligned} \tag{A8}$$

Each term in the decomposition of Eq. (A7) is an operator with two-dimensional support and these supports are all mutually orthogonal. Specifically the support of $\hat{\rho}_{(x)}$ is $\{\hat{U}_{\text{prep}}|\overline{x}\rangle, \hat{U}_{\text{prep}}|\overline{N-x}\rangle\}$ and the unitary nature of the preparatory gate results in the orthogonality of distinct supports.

It remains to determine an expression for $\hat{U}_{\text{prep}}|\overline{x}\rangle$ in terms of the computational basis (i.e., with bits strings consisting of

"0" and "1" terms rather than "+" or "−" terms). For a single qubit

$$|\overline{x_j}\rangle = \frac{1}{\sqrt{2}} \sum_{z_j=0}^{1} (i)^{z_j} (-1)^{x_j z_j} |z_j\rangle, \qquad (A9)$$

and this gives

$$|\overline{x}\rangle = \frac{1}{2^{n/2}} \sum_{z=0}^{N} (i)^{z_1+\dots+z_n} (-1)^{x \cdot z} |z\rangle. \qquad (A10)$$

Combining Eqs. (A3) and (A10) gives

$$\hat{U}_{\text{prep}}|\overline{x}\rangle = \frac{1}{2^n} \sum_{y,z=0}^{N} (i)^{z_1+\dots+z_n} (-1)^{y \cdot z + x \cdot z + s} |y\rangle \qquad (A11)$$

where $s := z_1 z_2 + z_1 z_3 + \dots + z_{n-1} z_n$, in which each distinct pair of nonequal subscripts appears exactly once. We shall show that

$$\hat{U}_{\text{prep}}|\overline{x}\rangle = \frac{1+i}{2} |x\rangle + \frac{1-i}{2} |N-x\rangle \qquad (A12)$$

(recall that $|x\rangle$ contains the bit string for $x$ in terms of "0" and "1" while $|\overline{x}\rangle$ contains the bit string for $x$ in terms of "+" and "−"). Substitution into Eq. (A8) then yields the result of Eq. (33).

In order to show Eq. (A12), consider $\langle x|\hat{U}_{\text{prep}}|\overline{x}\rangle$. Then Eq. (A11) gives

$$\langle x|\hat{U}_{\text{prep}}|\overline{x}\rangle = \frac{1}{2^n} \sum_{z=0}^{N} (i)^{z_1+\dots+z_n} (-1)^{x \cdot z + x \cdot z + s}$$

$$= \frac{1}{2^n} \sum_{z=0}^{N} (i)^{z_1+\dots+z_n} (-1)^{s} \qquad (A13)$$

since $(-1)^{x \cdot z + x \cdot z} = 1$. The term in the sum, $(i)^{z_1+\dots+z_n} (-1)^s$ depends only on the number, $m$, of $z_1, \dots, z_n$ which are equal to 1. Clearly $\gamma := (i)^{z_1+\dots+z_n} = i^m$. The only bits which contribute to $s$ are those for which *both* $z_j = 1$ and $z_k = 1$. Given that $m$ of $z_1, \dots, z_n$ are equal to 1, the number of such such pairs is $\binom{m}{2} = m(m-1)/2$. Depending on whether this is even or odd, the term $\beta := (-1)^s$ gives $\pm 1$. The various possibilities for $m$ and their contributions are summarized in Table II.

This shows that whenever $m$ is even the contribution to the sum is $\beta\gamma = 1$ and whenever it is odd, $\beta\gamma = i$. As $z$ ranges through all possible values, the number of times that $m$ is even is $2^n/2$ (exactly half of all bit strings have even numbers of bits equal to 1). Thus there are $2^n/2$ contributions of 1 and

TABLE II. Contributions to the sum of Eq. (A13). Here $k = 0, 1, \dots$ is an integer.

| $m$ | $\beta$ | $\gamma$ | $\beta\gamma$ |
|---|---|---|---|
| $4k$ | 1 | 1 | 1 |
| $4k+1$ | 1 | $i$ | $i$ |
| $4k+2$ | $-1$ | $-1$ | 1 |
| $4k+3$ | $-1$ | $-i$ | $i$ |

$2^n/2$ contributions of $i$. Thus

$$\langle x|\hat{U}_{\text{prep}}|\overline{x}\rangle = \frac{1+i}{2}. \qquad (A14)$$

Now consider $\langle N - x|\hat{U}_{\text{prep}}|\overline{x}\rangle$. Note that each bit in $\langle N - x|$ is the bit flip of the corresponding bit in $\langle x|$. Then in Eq. (A11)

$$\langle N - x|(-1)^{y \cdot z} |y\rangle = (-1)^{x \cdot z + z_1 + \dots + z_n}. \qquad (A15)$$

Thus

$$\langle N - x|\hat{U}_{\text{prep}}|\overline{x}\rangle = \frac{1}{2^n} \sum_{z=0}^{N} (-i)^{z_1+\dots+z_n} (-1)^{s} \qquad (A16)$$

which shows that $\langle N - x|\hat{U}_{\text{prep}}|\overline{x}\rangle = (\langle x|\hat{U}_{\text{prep}}|\overline{x}\rangle)^* = (1 - i)/2$. This determines the two coefficients that appear in Eq. (A12). Since $\hat{U}_{\text{prep}}|\overline{x}\rangle$ has unit modulus, there cannot be any other components. This demonstrates Eq. (A12) and proves Eq. (33).

## APPENDIX B: BOUNDS ON $G(\lambda)$

The general form of the gain is given by Eq. (42), which contains the term $c_j^2/d_j^2$. This enters into various bounds on the gain and following is a key result.

*Proposition 3.* Suppose that $r < 1$. If $j \neq n/2$ then

$$\frac{c_j^2}{d_j^2} \geqslant r^2 \qquad (B1)$$

with equality if and only if $r = 0$. If $j = n/2$ then

$$\frac{c_j^2}{d_j^2} = 0. \qquad (B2)$$

*Proof.* Clearly if $r = 0$, then $c_j = 0$ and thus $c_j/d_j = r$. Thus assume that $r > 0$. Let

$$f_j := \frac{c_j}{d_j} \frac{1}{r}. \qquad (B3)$$

By the definition of $c_j$ and $d_j$, it follows that $f_{n-j} = -f_j$. Thus it suffices to prove the result for $j \geqslant n/2$.

Two special cases present themselves. First, if $n$ is even and $j = n/2$, Eq. (40) gives that $c_j = 0$. Second, if $n$ is odd and $j = (n+1)/2$, then let $n = 2l + 1$. Thus $j = l + 1$ and direct substitution into Eq. (40) gives $c_j/d_j = r$.

For all other cases $j > (n+1)/2$. Here

$$f_j - 1 = \frac{d_j(1 - r) - 2(1 + r)^{n-j}(1 - r)^j}{d_j r}. \qquad (B4)$$

The numerator of this expression gives

$$d_j(1 - r) - 2(1 + r)^{n-j}(1 - r)^j$$
$$= (1 + r)^j(1 - r)^{n-j+1} + (1 + r)^{n-j}(1 - r)^{j+1}$$
$$\quad - 2(1 + r)^{n-j}(1 - r)^j$$
$$= (1 + r)^j(1 - r)^{n-j+1} - (1 + r)^{n-j+1}(1 - r)^{j+1}. \qquad (B5)$$

Now if $0 \leqslant r < 1$, then $[(1 + r)/(1 - r)]^{2j-1-n} \geqslant 1$ provided that $j > (n+1)/2$ (note that equality occurs if and only if $r = 0$). Thus $(1 + r)^{2j-1-n} \geqslant (1 - r)^{2j-1-n}$. This implies that $(1 + r)^j(1 - r)^{n-j+1} - (1 + r)^{n-j+1}(1 - r)^{j+1} \geqslant 0$, with equality if and only if $r = 0$.. Furthermore, $d_j > 0$ and for $r > 0$,

Eq. (B4) gives that $f_j - 1 > 0$. This proves the proposition. ∎

It follows from Proposition 3 that, if $r < 1$ then the summand in Eq. (42) satisfies

$$\frac{c_j^2}{d_j} \frac{1 - (1 - 2\lambda)^2 r^2}{1 - (1 - 2\lambda)^{2m} c_j^2 / d_j^2} \geqslant \frac{c_j^2}{d_j} \frac{1 - (1 - 2\lambda)^2 r^2}{1 - (1 - 2\lambda)^{2m} r^2} \quad (B6)$$

with equality only possible if $r = 0$ (note that for $j = n/2$, the result of Proposition 3 is irrelevant since $c_j = 0$). Thus

$$G(\lambda) \geqslant \frac{m(1 - 2\lambda)^{2m-2}}{2^{n+1} r^2} \frac{1 - (1 - 2\lambda)^2 r^2}{1 - (1 - 2\lambda)^{2m} r^2} \sum_{j=0}^{n} \binom{n}{j} \frac{c_j^2}{d_j}. \quad (B7)$$

In this context, the following is useful.

*Proposition 4.* If $n \geqslant 2$ then

$$\sum_{j=0}^{n} \binom{n}{j} \frac{c_j^2}{d_j} \geqslant 2^{n+1} r^2 \quad (B8)$$

with equality if $r = 0$.

*Proof.* If $r = 0$ then $c_j = 0$ and the equality holds. In general, straightforward algebra yields

$$c_j^2 = d_j^2 - 4(1 - r^2)^n. \quad (B9)$$

Thus

$$\sum_{j=0}^{n} \binom{n}{j} \frac{c_j^2}{d_j} = \sum_{j=0}^{n} \binom{n}{j} d_j - 4(1 - r^2)^n \sum_{j=0}^{n} \binom{n}{j} \frac{1}{d_j}$$

$$= 2^{n+1} - 4(1 - r^2)^n \sum_{j=0}^{n} \binom{n}{j} \frac{1}{d_j}. \quad (B10)$$

It will be shown that

$$d_j \geqslant 2(1 - r^2)^{n-1} \quad (B11)$$

and thus

$$4(1 - r^2)^n \sum_{j=0}^{n} \binom{n}{j} \frac{1}{d_j} \leqslant 2(1 - r^2) \sum_{j=0}^{n} \binom{n}{j}$$

$$= 2^{n+1}(1 - r^2). \quad (B12)$$

This will establishes the proposition.

It remains to prove Eq. (B11). It is only necessary to show this for $j \geqslant n/2$ since $d_{n-j} = d_j$. First, for $n = 2$, there are two cases. With $j = 1$, $d_1 = 2(1 - r^2)$ and with $j = 2$, $d_2 = 2(1 + r^2)$. Equation (B11) is clearly satisfied for both cases. Suppose that $n > 2$ and let

$$\phi(r) := 2\frac{(1 - r^2)^{n-1}}{d_j.} \quad (B13)$$

Then if $j = n/2$, $d_j = 2(1 - r^2)^{n/2}$ and $\phi(r) = (1 - r^2)^{n/2-1} \leqslant 1$ with equality if and only if $r = 0$. Suppose that $j > (n + 1)/2$. Then we shall show that $\phi(r)$ is a monotonically decreasing function of $r$ for $0 < r < 1$. First, $\phi(0) = 1$ and $\lim_{r \to 1} \phi(r) = 0$. Now consider

$$\frac{d\phi}{dr} = \frac{(1 - r^2)^{n-2}}{d_j^2} [r(2 - n)d_j + (n - 2j)c_j]. \quad (B14)$$

This is a continuous function of $r$ and if $\phi(r)$ is not monotonically decreasing or monotonically increasing, then $\frac{d\phi}{dr}$ must equal zero at some point in the range $0 < r < 1$. Then $\frac{d\phi}{dr} = 0$ if and only if $r(2 - n)d_j + (n - 2j)c_j$. This is equivalent to $r(2 - n)(1 + \alpha^{2j-n}) = (2j - n)(1 - \alpha^{2j-n})$ where $\alpha = (1 - r)/(1 + r)$. Clearly $\alpha < 1$. But if $n > 2$ and $j > n/2$ then the left-hand side is strictly less than 0 while the right is positive or 0. This is impossible and $\frac{\partial\phi}{\partial r} \neq 0$. Thus $\frac{d\phi}{dr} \neq 0$ and has the same sign throughout the range $0 < r < 1$. So $\phi(r)$ is either monotonically decreasing or monotonically increasing. The latter is true since $\phi(1) < \phi(0)$. This shows that $\phi(r) < \phi(0) = 1$ and proves Eq. (B11). ∎

The results of Proposition 4 together with Eq. (B7) imply that for $0 < r < 1$,

$$G(\lambda) > m(1 - 2\lambda)^{2m-2} \frac{1 - (1 - 2\lambda)^2 r^2}{1 - (1 - 2\lambda)^{2m} r^2}. \quad (B15)$$

It follows that with a single invocation of the phase-flip operation (i.e., $m = 1$), $G(\lambda) > 1$ whenever $0 < r < 1$.

### APPENDIX C: PROPERTIES OF $\mathcal{Q}$

Straightforward differentiation and algebra shows that

$$\frac{\partial \mathcal{Q}}{\partial \mu} > 0 \quad (C1)$$

whenever $\mu > 0$.

If $r \geqslant \mu > 0$ then

$$\frac{\partial \mathcal{Q}}{\partial r} = \frac{r}{2} \log_2 \left[ \frac{(1 + r^2 + 2r\mu)(1 + r^2 - 2r\mu)}{(1 + r^2)^2} \right]$$

$$+ \frac{\mu}{2} \log_2 \left[ \frac{(1 + r^2 + 2r\mu)}{(1 + r^2 - 2r\mu)} \right]$$

$$\geqslant \frac{\mu}{2} \log_2 \left[ \frac{(1 + r^2 + 2r\mu)(1 + r^2 - 2r\mu)}{(1 + r^2)^2} \right]$$

$$+ \frac{\mu}{2} \log_2 \left[ \frac{(1 + r^2 + 2r\mu)}{(1 + r^2 - 2r\mu)} \right]$$

$$= \frac{\mu}{2} \log_2 \left[ \frac{(1 + r^2 + 2r\mu)^2}{(1 + r^2)^2} \right]. \quad (C2)$$

The argument of this logarithm is larger than or equal to 1 (only possible if $r = 0$ or $\mu = 0$). This shows that if $r > \mu$ then $\frac{\partial \mathcal{Q}}{\partial r} > 0$.

On the other hand if $\mu > 0$ and $r \leqslant \mu$ then

$$\frac{\partial \mathcal{Q}}{\partial r} = \frac{r}{2} \log_2 \left[ \frac{(1 + r^2 + 2r\mu)(1 + r^2 - 2r\mu)}{(1 + r^2)^2} \right]$$

$$+ \frac{\mu}{2} \log_2 \left[ \frac{(1 + r^2 + 2r\mu)(1 - r\mu)}{(1 + r^2 - 2r\mu)(1 + r\mu)} \right]$$

$$\geqslant \frac{r}{2} \log_2 \left[ \frac{(1 + r^2 + 2r\mu)^2(1 - r\mu)}{(1 - r^2)^2(1 + r\mu)} \right]. \quad (C3)$$

Let

$$\omega := (1 + r^2 + 2r\mu)^2(1 - r\mu) - (1 - r^2)^2(1 + r\mu). \quad (C4)$$

Then

$$\omega = 2r[\mu(1 - r^2)(1 + r^2) + 2r(1 - r^2\mu^2) + 2r^2\mu(1 - \mu^2)] \tag{C5}$$

and for every $0 \leqslant r, \mu \leqslant 1$ every term here is positive. Thus $\omega > 0$ and the argument of the last logarithm is larger than 1. Thus if $\mu > 0$ and $r \leqslant \mu$ then $\frac{\partial \mathcal{Q}}{\partial r} > 0$. This shows that in all cases $\frac{\partial \mathcal{Q}}{\partial r} > 0$.

---

[1] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[2] C. M. Caves, K. S. Thorne, R. W. P. Drever, V. D. Sandberg, and M. Zimmermann, Rev. Mod. Phys. **52**, 341 (1980).

[3] J. H. Shapiro and S. R. Shepard, Phys. Rev. A **43**, 3795 (1991).

[4] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).

[5] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439 (1994).

[6] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).

[7] M. Sarovar and G. J. Milburn, J. Phys. A **39**, 8487 (2006).

[8] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).

[9] M. Paris, Int. J. Quantum Inf. **9**, 125 (2009).

[10] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, Phys. Rev. A **80**, 052114 (2009).

[11] C. O'Loan, e-print arXiv:1001.3971v1 [quant-th].

[12] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, Nat. Phys. **7**, 406 (2011).

[13] V. Giovannetti, S. Lloyd, and L. Maccone, Nature Photonics **5**, 222 (2011).

[14] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[15] C. P. Slichter, *Principles of Magnetic Resonance* (Springer, Berlin, 1996).

[16] J. A. Jones, Prog. NMR Spectrosc. **59**, 91 (2011).

[17] M. Zwierz, C. A. Pérez-Delgado, and P. Kok, Phys. Rev. Lett. **105**, 180402 (2010).

[18] M. Zwierz, C. A. Pérez-Delgado, and P. Kok, Phys. Rev. A **85**, 042112 (2012).

[19] A. Fujiwara and H. Imai, J. Phys. A **36**, 8093 (2003).

[20] C. J. O'Loan, J. Phys. A **40**, 14499 (2007).

[21] Z. Ji, G. Wang, R. Duan, Y. Feng, and M. Ying, IEEE Trans. Inf. Theory **54**, 5172 (2008).

[22] A. Chiuri, V. Rosati, G. Vallone, S. Pádua, H. Imai, S. Giacomini, C. Macchiavello, and P. Mataloni, Phys. Rev. Lett. **107**, 253602 (2011).

[23] L. Ruppert, D. Virosztek, and K. Hangos, J. Phys. A **45**, 265305 (2012).

[24] G. M. D'Ariano, C. Macchiavello, and P. Perinotti, Phys. Rev. A **72**, 042327 (2005).

[25] S. Boixo and R. D. Somma, Phys. Rev. A **77**, 052320 (2008).

[26] K. Modi, H. Cable, M. Williamson, and V. Vedral, Phys. Rev. X **1**, 021022 (2011).

[27] A. Datta, L. Zhang, N. Thomas-Peter, U. Dorner, B. J. Smith, and I. A. Walmsley, Phys. Rev. A **83**, 063836 (2011).

[28] W. van Dam, G. M. D'Ariano, A. Ekert, C. Macchiavello, and M. Mosca, Phys. Rev. Lett. **98**, 090501 (2007).

[29] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946).

[30] O. E. Barndorff-Nielsen and R. D. Gill, J. Phys. A: Math. Gen. **33**, 4481 (2000).

[31] D. Collins, K. W. Kim, W. C. Holton, H. Sierzputowska-Gracz, and E. O. Stejskal, Phys. Rev. A **62**, 022304 (2000).

[32] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).

[33] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996).

[34] H. Ollivier and W. H. Zurek, Phys. Rev. Lett. **88**, 017901 (2001).

[35] K. Modi, A. Brodutch, H. Cable, T. Paterek, and V. Vedral, Rev. Mod. Phys. **84**, 1655 (2012).

[36] S. Luo, Phys. Rev. A **77**, 042303 (2008).

[37] Q. Chen, C. Zhang, S. Yu, X. X. Yi, and C. H. Oh, Phys. Rev. A **84**, 042313 (2011).

[38] M. Gu, H. M. Chrzanowski, S. M. Assad, T. Symul, K. Modi, T. C. Ralph, V. Vedral, and P. K. Lam, Nat. Phys. **8**, 671 (2012).

[39] V. Madhok and A. Datta, Int. J. Mod. Phys. B **27**, 1345041 (2013).

[40] L. J. Schulman, T. Mor, and Y. Weinstein, Phys. Rev. Lett. **94**, 120501 (2005).

[41] J. Baugh, O. Moussa, C. A. Ryan, R. Laflamme, C. Ramanathan, T. F. Havel, and D. G. Cory, Phys. Rev. A **73**, 022305 (2006).

[42] R. Marx, A. F. Fahmy, J. M. Myers, W. Bermel, and S. J. Glaser, Phys. Rev. A **62**, 012310 (2000).

[43] R. Das, T. Mahesh, and A. Kumar, J. Mag. Res. **159**, 46 (2002).

[44] X. Peng, X. Zhu, X. Fang, M. Feng, M. Liu, and K. Gao, Phys. Rev. A **65**, 042315 (2002).

[45] C. Ramanathan, N. Boulant, Z. Chen, D. G. Cory, I. Chuang, and M. Steffen, Quantum Inf. Proc. **3**, 15 (2004).

[46] L. Xiao and J. A. Jones, Phys. Rev. A **72**, 032326 (2005).